Richard F. Meraz

3009 Dana St. #C, Berkeley, CA, 94705 • 510-486-0118
rfmeraz@lbl.gov • http://frna.lbl.gov/

OBJECTIVE

To develop computational methods and software tools for solving challenging problems in Biology or other fields.

SKILLS

Computers and Computation: small and medium-scale software engineering, algorithm design, relational database design, web-software enginnering, linux system administration, machine-learning methods and statistical analysis.

Communication: writing articles and tutorials, grant-writing, teaching, world wide web, documenting software.

Bioinformatics: probabilistic analysis of biological sequences, machine learning approaches to sequence analysis, standard and non-standard bioinformatics tools.

EXPERIENCE HIGHLIGHTS

Computer programming and Bioinformatics

Languages and development: Perl, C, Java, Bash, Python, Emacs, SQL (MySQL), Apache and mod-perl, CGI, CVS, LaTeX

Operating Systems and Environments: GNU/Linux, Unix, Microsoft Windows, cluster environments (openPBS and SUN Grid Engine)

Mathematical Methods: Machine learning methods including support vector machines, neural networks, and cluster analysis. GNU Scientific library, Statistical Analysis, mathematical modeling.

Bioinformatics: BioPerl. Standard and probabalistic methods for biologial sequence analysis including: BLAST-varieties (WU, NCBI), HMMer, Stochastic Context Free Grammars (Infernal/ Covariance Models of RNA), CLUSTALW/TCoffee/MAFFT, EMBOSS (GCG). Biological Information Databases and Portals including NCBI-ENTREZ, SWISSPROT, ENSEMBL, MIPS, TIGR, KEGG, GMOD, PubMed, Gene Ontology. Dynamic programming algorithms for RNA and protein sequence analysis. Familiar with most general bioinformatics tools and many exotic and unusual ones.

Research

Learning RNA secondary structure from a dual graph representation: Support Vector Machines were used with marginalized kernels defined on general graph structures to learn the secondary structures of a variety of RNA families. This method is applicable to class specific RNA genefinding and riboswitch identification in genomic sequence as well as class discovery in sets of unannotated RNA sequence. To our knowledge this is the first application of discriminative learning using kernels defined directly on the topology of RNA secondary structure. July 2003 - January 2004

A bipartite-graph model of protein-complex data: Biological interpretation of results from unsupervised machine learning (min-max-cut clustering) of bipartite-graph model of protein-complex datasets. Applied computable vocabularies of the Gene Ontology to determine and interpret the biological context of statistically significant modules in the interaction network. Jan 2003 - July 2003

Applications of support vector machines to RNA gene-finding: Modeled genomic sequence signal for RNA-Genes in *C. elegans* and *E. coli* using physical parametrization of sequence and comparative genomics information. An algorithm for learning on partially labelled data using Support Vector Machines was used to train classifiers for ab-initio prediction of RNA-genes. Co-wrote a NIH grant (3 years 750K) with Dr. Stephen Holbrook for further work on these methods which was funded in February 2003. Sept 2001 - present

Neural networks and support-vector-machines for predicting protein function from sequence: Neural networks were applied to the prediction of protein functional categories from parametrizations of primary sequence alone. Applied method to determine nucleic-acid binding proteins for preliminary characterization of mechanisms for radiation-resistance in *D. radiodurans*. Jun 2000 - Jan 2001

Primary sequence analysis of metallothionein: Purified cadmium induced metallothionein from *L. littorea* and performed analysis with MALDI-MS at University of Odense in Denmark. May 1999 - Sept 1999

Metal metabolism in invertebrates: Worked to elucidate the cellular and molecular basis for the bio-accumulation, bio-amplification and detoxification of trace-metals and pollutants in marine organisms. Assisted in the development of a system for sequentially coupled SE/IE HPLC and Inductive Coupled Plasma Mass Spectrometry for measurement of trace-metal flux in cellular and physiological compartments. Jan 1996 - Apr 1999

Teaching

Using Perl for automation in a laboratory environment: Designed and taught a course at Lawrence Berkeley National Laboratory that illustrated Perl Programming techniques appropriate for automation of data analysis in a laboratory environment. Summer 2002

EMPLOYMENT HISTORY

Scientific Programmer II	Lawrence Berkeley National Laboratory	June 2003 - Dec 2003
DOE Fellow/Scientist	Lawrence Berkeley National Laboratory	June 2000 - June 2003
NIH (MARC) Fellow	Molecular Toxicology Laboratory, CSULB	Jan 1996 - Apr 1999
Teacher	The Princeton Review	Jan 1998 - March 1999

EDUCATION

B.S. Mathematics (minor Biochemistry)	California State University Long Beach	2000
Various Coursework	University of California, Berkeley	2001-present

PUBLICATIONS

Karklin, Y., Meraz R.F., Holbrook, S.R. (Submitted) The Shape of RNA: Can Secondary Structure Alone Predict Family. **Intelligent Systems in Molecular Biology**

Ding, C.H.Q., He, Xiaofeng, Meraz, R.F., Holbrook, S.R. (In Press) A Unified Representation of Multi-protein Complex Data for Analyzing Interaction Networks. **Proteins: Structure Function and Bioinformatics**.

Meraz, R.F., Carter R.J., Holbrook, S.R. (2003) Computational Methods for RNA Genefinding. **Nature Encylopedia of the Human Genome**.

Mason, A.Z., Meraz R.F. (1998) Cytosolic Metal Speciation Studies by Sequential, On-line, SE-IE/HPLC-ICPMS. **Marine Environmental Research**

Meraz, R.F., Holbrook, S.R. (In Preparation) Using Support Vector Machines to Identify RNA-Genes in *E. coli*.

Mason, A.Z., Meraz R.F. (In Preparation) A model of copper metabolism in the marine prosobranch mollusc *Littorina littorea*.

REFERENCES

Stephen R. Holbrook (srholbrook@lbl.gov)

Andrew Z. Mason (zedmason@csulb.edu)

Chris H. Q. Ding (chqding@lbl.gov)